如何建立聊天机器人与用户间的数字治疗联盟：关系线索的作用*

磨 然[1] 方作之[4#] 方建东[1,2,3]

[1](广西师范大学教育学部心理学系，桂林 541006)
[2](广西师范大学认知神经科学与应用心理广西高校重点实验室，桂林 541006)
[3](广西高校人文社会科学重点研究基地广西民族教育发展研究中心，桂林 541006)
[4](陕西师范大学心理学院，西安 710000)

摘 要：近年来，研究者们将“治疗联盟”(TA)的概念与在线自助干预(ISIs)相结合，以解决ISIs中用户参与度较低的问题。这种在数字环境中形成的TA，被称之为“数字治疗联盟”(DTA)。随着人工智能的迅速发展，聊天机器人可模拟人类指导，相对于传统ISIs程序更易于与用户建立关系，可通过友好、尊重、倾听、鼓励、真诚、理解、信任这几个关系线索来促进DTA的发展，为解决用户低参与度的问题提供了一种新思路。未来的研究可从影响因素、ISIs技术迭代、测量规范、实验操纵等方面对DTA作进一步的探索。

关键词：数字治疗联盟，聊天机器人，关系线索
分类号：B849

How to establish a digital therapeutic alliance between chatbots and users: The role of relational cues

MO Ran[1], FANG zuozhi[4#], FANG Jiandong[1,2,3]

[1](*Department of Psychology, Faculty of Education, Guangxi Normal University, Guilin* 541006, China)
[2](*Guangxi University and College Key Laboratory of Cognitive Neuroscience and Applied Psychology, Guangxi Normal University, Guilin* 541006, China)
[3](*Guangxi Ethnic Education Development Research Center, Key research base of Humanities and Social Sciences in Guangxi Universities, Guilin* 541006, China)
[4](*School of Psychology, Shaanxi Normal University, Xi'an* 710062, China)

Abstract: To address the issue of users' poor engagement, researchers have recently integrated the therapeutic alliance (TA) concept with Internet-based self-help interventions (ISIs). Digital therapeutic alliance (DTA) are TAs established within a digital environment. A chatbot can replicate human guidance due to the rapid development of artificial intelligence, and it is easier to establish relationships with users than traditional ISIs. Furthermore, it may enhance DTA through amiability,

---

收稿日期：2022-06-11
* 广西哲学社会科学规划研究课题、广西教育科学“十四五”规划2022年度重点委托课题 (2022AA11)、广西民族教育发展研究中心2022年度课题(2022MJZD001)、2022 年广西研究生教育创新计划项目(JGY2022059)资助。
#为并列第一作者。
通信作者：方建东，E-mail: bdjs2004@163.com

respectfulness, attentiveness, encouragement, sincere comprehension, and mutual trust, which presents a novel solution to this issue. Future research can investigate DTA from the perspectives of affecting factors, technology iteration of ISIs, measurement specification, and experimental manipulation.
Keywords: digital therapeutic alliance, chatbot, relational cues

1 引言

目前，在线自助干预(Internet-based Self-help Interventions, ISIs)的可行、有效性已得到广泛验证(Izzaty et al., 2021; Johansson et al., 2021; Sun et al., 2021; Taylor et al., 2021; Weisel et al., 2019)，或可成为面对面治疗的有力补充(Berry et al., 2019)，但高脱落率、低参与度仍是其众所周知的困境，对于无指导 ISI，此问题则更为凸显(Linardon et al., 2019; Pratap et al., 2020; Zhang et al., 2021)。随着人工智能(Artificial Intelligence, AI)技术的迅猛发展，能模拟人类对话的聊天机器人(Chatbot)可让无指导 ISI 在自动化后兼顾效率及成本效益(Luo et al., 2022)。具体而言，在聊天机器人的设计中引入关系线索(Relational Cues)，如自我披露、真诚、理解、幽默等(Gallen et al., 2018)，可在认知、情感两个维度上满足用户的需要(Abd-alrazaq et al., 2019; Provoost et al., 2017; Wiese et al., 2022)，并与用户建立数字治疗联盟(Digital Therapeutic Alliance, DTA)，进而促进用户的参与度及治疗效果(Goldberg et al., 2021; Liu et al., 2022; Provoost, 2021)。纵观既有研究，国外相关成果尽管丰富但较为零散，而国内关于 ISIs 的研究进展尚处初期阶段(Henson et al., 2019; Grekin et al., 2019; Yao et al., 2020; Zhang et al., 2021)。鉴于此，文章集中探讨聊天机器人在无指导 ISI 中通过关系线索对 DTA 产生的可能影响，以引发同行们对该领域的研究兴趣，为进一步的研究提供参考。

2 DTA 的发展

治疗联盟(Therapeutic Alliance, TA)，也称工作同盟(Working Alliance, WA)，是来访者和咨询师之间为实现治疗目标而合作的关系的质量与强度(朱旭, 江光荣, 2011a)。上世纪 70 年代，Bordin(1979)将 TA 分为三个成分——情感纽带、对治疗任务达成共识、就治疗目标达成一致，并成为 TA 最流行之定义。而后，Horvath 与 Greenberg(1989)基于 Bordin 的定义编制了首个 TA 量表——工作同盟量表(Working Alliance Inventory, WAI)。近年来，随着 WAI 开始逐渐被用于数字心理健康的研究(Andersson et al., 2012)，DTA 一词也由此诞生。在诸如以电子邮件、在线聊天、视频会议、ISIs 程序等干预形式所建立的 TA 均可称为 DTA(D’Alfonso et al., 2020; Henson et al., 2019; Lederman & D’Alfonso, 2021)。

DTA 之所以得到发展，可能的原因有三：第一，新冠疫情加速了社会的数字化进程，虚拟现实(Virtual Reality)、元宇宙(Metaverse)等概念也得到了发展(张夏恒, 李想, 2022)。从哲学的创世观看，人类虽生活在既定的宇宙中且被排除于创世者之外，但人类一直有超越自然的梦想，而数字化的发展则为人类提供了造世的机会(黄欣荣,曹贤平, 2022)。因此，未来现实生活的数字化将成为客观趋势，人机关系也变得越来越重要。第二，ISIs 正朝着效益最大化的方向发展，但更高水平的自动化也伴随着用户参与度低、脱落率高的问题，基于此，在面对面心理咨询/心理治疗中占有重要地位的 TA 也自然受到研究者们的关注。第三，基于自我决定理论(Self-determination Theory,

SDT)，自主、胜任、关系这三个基本需求的满足，能促进个体的外在动机向内在动机转化，进而保证其心理健康成长(Deci & Ryan, 1985)。与之相应，ISIs 程序可辅助用户解决问题，提高用户的自主感、胜任感，进而有助于培养 TA 中的情感纽带。同时，TA 中“在目标和任务上达成一致”与用户使用 ISIs 程序时的目标确立及所获得的阶段性反馈有关。因此，DTA 在可行性、有效性方面具有一定的理论支撑。

有一系列研究表明，在 ISIs 中建立的 DTA 与面对面心理咨询/心理治疗中建立的 TA 水平相接近(Andersson et al., 2012; Heim et al., 2018; Klasen et al., 2013; Pihlaja et al., 2018; Tremain et al., 2020)。同时，DTA 与参与度呈正相关(Baumel & Kane, 2018; Goldberg et al., 2021; Hargreaves et al., 2018; Heim et al., 2018; Perski et al., 2017; Rodrigues et al., 2021)，而参与度则是改善 ISIs 治疗效果的关键因素(Arndt et al., 2020; Asaeikheybari et al., 2021; Fuhr et al., 2018; Puls et al., 2020)。另一项元分析还指出，DTA 与治疗效果也存在相关，且总体效应量中等，但实际的研究结果好坏参半(Probst et al., 2019)，这与测量工具的发展及选用不无关系。DTA 测量需针对数字环境具体考量，若简单改编传统 WAI，可能无法解释数字干预中 TA 的复杂性。研究者们逐渐认识到这一点，并着手将传统 WAI 基于数字环境进行改编(Tremain et al., 2020)。例如，Berger 等人(2014)在 WAI-SR 的基础上进行改编，使之适应有指导 ISI。最初，Kiluk 等人(2014)提出基于原版 WAI 的数字改编版(WAI-Tech)，用于无指导 ISI 中 DTA 的测量，但其仅仅只将量表中的“咨询师”换成了“应用程序”。随着对 DTA 的进一步探究，Meyer 等人(2015)在研究中发现，被试和 ISIs 程序之间的 TA 并不等同于和人类咨询师之间的 TA。因此，他们对帮助联盟问卷(Helping Alliance Questionnaire, HAQ)进行了改编，以评估被试在多大程度上认为程序有所助益。在实证研究中，被试在干预后第 3 周的 HAQ 得分即可成功预测其 3 个月后的治疗效果。最近，Berry 等人(2018)考虑了无指导 ISI 的特点，在 阿格纽关系量表(Agnew Relationship Measure, ARM)的基础上编制了移动版阿格纽关系量表(Mobile Agnew Relationship Measure, mARM)。随后，Henson 等人(2019)在 WAI-SR 的基础上编制了 D-WAI，以专门评估无指导 ISI 的 DTA。Gómez Penedo 等人(2020)也为更好地测量有指导 ISI 中的 DTA，在 Berger(2014)的基础上编制了 WAI-I，并在大样本中验证了此量表的可靠性。为进一步使 DTA 测量适应数字干预场景，D’Alfonso 等人(2020)在 mARM 的基础上，将人机交互(Human-Computer Interaction, HCI)理论与 TA 理论结合，并尝试开发一种能更可靠地在无指导 ISI 中评估 DTA 的量表。时下，聊天机器人技术正不断地改变传统无指导 ISI 程序的交互体验，其既提供了类似于人类的指导，但又实现了全自动化。因此，未来 DTA 测量的发展或将与新兴 AI 技术的迭代趋势相适应。

3 关系线索或是影响 DTA 的重要因素

目前，ISIs 多是基于认知行为疗法(Cognitive Behavioral Therapy, CBT)进行设计，且对压力、抑郁、焦虑、烟瘾、酒瘾、失眠及创伤后应激障碍等问题均有显著的疗效(Weisel et al., 2019)。根据 Bielinski 和 Berger(2020)的划分，ISIs 的常见类型有三：一是无指导干预(Unguided Interventions)，指在线干预的过程中无咨询师介入，用户仅通过程序自助；二是有指导干预(Guided Interventions)，指将用户自助与定期、简短的在线辅导(同步或异步)相结合；三是混合干预(Blended Interventions)，指将在线干预与面对面心理咨询/心理治疗相结合，以前者作为后者的补充。

在 ISIs 的情境中，若有咨询师的支持，TA 则相对更容易建立。有研究指出，有指

导 ISI 与面对面治疗所建立的 TA 水平并无显著差异，TA 不但预测了参与度，也预测了治疗效果(Anderson et al., 2012; Kaiser et al., 2021)。尽管有指导 ISI 的整体效果往往优于无指导 ISI(Baumeister et al., 2014)，但也有研究表明，在低强度的有指导 ISI 中，被试在干预初期的情感纽带得分较低且增速缓慢(Jasper et al., 2014)，而被试在参与度、治疗效果上的得分也与无指导 ISI 上的对应得分无显著差异(Chen et al., 2020)，这说明人类介入的缺乏限制了 TA 的发展。不过，Berry 等人(2018)指出，被试实际上也能与传统的无指导 ISI 程序发展虚拟关系，这有助于弥补缺乏人类指导带来的影响。Holter 等人(2020)以扎根理论(Grounded Theory)建立的人机关系模型也认为，个体与无指导 ISI 程序能够建立情感纽带，但前提是要使个体对程序的感知在社会行动者与无生命的程序之间交替转换。基于此，研究者们开始尝试在无指导 ISI 程序中加入基于传统编程的虚拟化身(Avatar)，以缩小其与有指导 ISI 效果的差距。例如，在 Heim 等人(2018)的研究中，被试的情感纽带得分因虚拟化身的加入而稳定发展，并与失眠改善相关。但是，一些被试却表示他们更想与人类咨询师交流，此意愿也预测了疗效。类似的，Fenski 等人(2021)指出，若虚拟化身不能对被试的负面情绪准确识别并给予恰当回应，则很有可能起到反作用。总的来说，嵌入于无指导 ISI 程序中的虚拟化身有希望与人类建立类似于有指导 ISI 中的 DTA，且 DTA 可正向影响参与度及治疗效果，但如何设计虚拟化身以保障疗效仍需进一步讨论。

人类线索(Human Cues)，是计算机程序因模拟人类形象、言语、行为等条件而具有的特征，能让与之交互的个体产生往往只有与真人交互时才特有的感受(Rodrigues et al., 2021)。社会行动者范式(Computers as Social Actors, CASA)也指出，人类往往会下意识地对计算机程序呈现的人类线索做出反应，且无论这些线索有多么初级(Nass et al., 1994)。为将人类线索具体化，Gallen 等人(2018)将其分为 4 类：视觉线索(Visual Cues)，如年龄、性别、外貌、表情、动作等；言语线索(Verbal Cues)，如文字、语音、语调、语速等；准非言语线索(Quasi-Nonverbal Cues)，如表情符号；关系线索(Relational Cues)，如自我披露、理解、幽默等。这些线索都可能对情感纽带及整个 DTA 的建立、发展造成影响，可作为指导虚拟化身设计的起点。

由上文可知，被试往往对虚拟化身存有更高的情感期待，而这种情感期待是否能得到满足也会影响情感纽带的发展。若情感纽带建立的不够牢固，则可能会限制 DTA 的发展，进而导致较差的参与度及治疗效果。然而，要形成情感纽带，前提则是虚拟化身向被试传递了温暖、安全和信任等关系线索(Negri et al., 2019)。在早期的研究中，研究者就已发现在虚拟化身的对话设计中引入寒暄、幽默、同理心等关系线索对情感纽带的影响相对目标、任务维度更大(Bickmore et al., 2005)。在最近的研究中，ter Stal 等人(2020)也指出，富有同理心的话语是影响人机关系的关键因素。因此，赋予虚拟化身恰当的关系线索对其与用户发展 DTA 可能有重要的作用。

4 如何设计关系线索来促进 DTA 的发展

若关系线索对 DTA 的发展可能起到重要的作用，那么，如何设计关系线索，并让其更高效地介入自然也变得重要(Mu, 2021)。此时，基于 AI 自然语言处理(Natural Language Processing, NLP)技术的聊天机器人就展现出了优势，其不但能呈现丰富的人类线索，还能基于用户的行为数据进行持续的“学习”(Zhang et al., 2020)，并给予用户个性化的反馈(Laranjo et al., 2018; Zhang et al., 2020)，比基于传统编程的虚拟化身更高效、灵活且人性化，无疑是更活跃的社会行动者(Alkhaldi et al., 2016; Ames et al., 2019; Hardeman et al., 2019; Tremain et al., 2020)。

自 1966 年世界上第一个真正意义上的聊天机器人 ELIZA 诞生以来(Weizenbaum,

1983)，聊天机器人的技术就一直在持续迭代，并逐步融入到数字心理健康之中(Elmasri & Maeder, 2016; Fitzpatrick et al., 2017; Gaffney et al., 2014)。目前，聊天机器人通常作为单独的功能模块嵌入 ISIs 程序之中，以语音用户界面(Voice User Interface, VUI)的形式为用户提供帮助，可替代人类咨询师的指导而使程序成为一种新型的无指导 ISI(如 MYLO, Woebot)，或是配合人类咨询师作为一个辅助的功能(如 dll 心聆“小天”)。此外，根据回复生成机制，聊天机器人可分为两类：一是检索式(Retrieval-based)，聊天机器人将从静态的知识库中检索预定义的规则来进行回复；二是生成式(Generation-based)，聊天机器人将通过学习及推理机制来动态生成回复(Song et al., 2018)。在形态方面，聊天机器人还可大致分为两类：一是具有虚拟化身，这是一种将聊天机器人和交互式化身(计算机生成的数字角色，其外观可能为人类或卡通人物)结合在一起的程序形态，可通过眼神、表情、动作、语音、文本等方式与人类交互(如 Replika)；二是仅以语音、文本与人类交互(如 Siri, 微软“小冰”)。

近年来，聊天机器人在 ISIs 中的应用逐渐增多，有研究发现其不但比传统的无指导 ISI 程序更能促进被试的参与度(Perski et al., 2019; Vaidyam et al., 2019)，且其与被试所建立的 TA 水平也与人类相当(Darcy et al., 2021)。尽管如此，若要问聊天机器人为何有效，研究者们却知之甚少。本文假设，可能的原因有四：第一，心智感知理论(Mind Perception Theory; Waytz et al., 2010)指出，个体可感知到其他对象具有心理能力，并对其作拟人化的信息加工。因此，聊天机器人的关系线索越丰富，就越可能提升社会存在(Social Presence)，使个体产生与真实人类交互的感知(Lee et al., 2020; Sundar, 2008)。同时，拟人化的聊天机器人通常比人类更可靠、易得，个体与其交互也容易获得更多安全感(Wanser et al., 2019) ，从而更倾向与其合作(Wiese et al., 2022)。第二，基于社会线索减少理论(Reduced Social Cues, RSC)，在网络文本信息交互的过程中，由于思想和情感必须转化为文字以弥补非言语信息的缺乏(Kiesler et al., 1984)。因此，个体在信息加工的过程中可能会产生网络去抑制效应(Online Disinhibition Effect, ODE)，进而表现出不同于面对面交流时的行为，包括放松、较少的约束感以及较开放的情感表达等(Suler, 2004)，这可能会使人机关系变得更为紧密、牢固。第三，聊天机器人天然具有人类线索，能拟人化地辅助个体解决问题，满足了 SDT 原则，进而促进情感纽带的发展。第四，由人际投资模型(The Investment Model of Personal Relationships)可知，聊天机器人提供的情感支持及有价值的信息，可使个体的感知获益及感知投入持续增加、认知成本及疑虑持续降低，进而逐渐建立信任感，提升对 ISIs 程序的参与度(Rusbult et al., 1994)。

归纳上述，具备关系线索且更为灵活的聊天机器人更利于从认知及情感的角度切入，在无指导 ISI 中促进 DTA 的快速发展，解决用户参与度低的问题。然而，尽管已有少部分研究者针对此问题进行了探索，但目前尚未有研究者归纳出确实、有效的关系线索以指导聊天机器人的设计。例如，Rodrigues 等人(2021)发现，与仅具有视觉线索的聊天机器人相比，仅具有关系线索的聊天机器人更能与被试建立 DTA，且参与度也更高。但是，此研究只探讨了视觉线索与关系线索在 DTA 上的差异，而并未检验不同关系线索对 DTA 的影响。为将聊天机器人的作用具体化，应对可能更为关键的关系线索作深入的探讨。因此，下文将在前人研究的基础之上提出几种 ISIs 中可能会对 DTA 带来积极影响的关系线索(Bordin, 1979; Horvath & Greenberg, 1989; Norcross, 2002)，以帮助聊天机器人发展人工智慧(Artificial Wisdom)。

4.1 友好尊重

在面对面咨询中，温暖、和谐、宽松、自由且安全的谈话氛围以及咨访双方的相互尊重都是 TA 的助长因素(Luborsky, 1976)。同样，在过往的 ISIs 研究中，友好与尊重也被认为是程序中必备的基础性设计，其中 ISIs 程序所传递信息的语气、语调都会对干预的可信度、参与度、有效性造成影响(Ames et al., 2019; Bock et al., 2015)。尽管被试的偏好各不相同，但礼貌、尊重、友好、幽默、积极等友好的对话语气相对更受其青睐，反之，被试普遍对压力、教训、羞辱等较为排斥(Ames et al., 2019; Mu, 2021)。在使用聊天机器人时，这种影响还可能被实时的对话交流强化。例如，当聊天机器人直呼被试的名字，并在适当的时候使用幽默，也能增进双方的友好关系(Bickmore et al., 2009)。原因在于，聊天机器人的“人格”特征会影响被试情绪反应的强度(Medhi Thies et al., 2017)，若被试因聊天机器人的互动反馈而将其归因为礼貌、友好、尊重的，即便被试知晓这是虚拟交互，也仍会将这种类社会互动(Parasocial Interaction)视作一种亲密的社交互动(Horton & Wohl, 1956)，并将对应的社会化规范应用于与聊天机器人的交互中。而当聊天机器人具有虚拟形象时，由于其呈现出更为丰富的非言语信息(如表情、姿势、动作、唇同步等)，还可通过人工情绪传染(Artificial Emotional Contagion)的机制(如模仿及情绪镜像)，使被试更容易感受到其友好、尊重之特征(Nofz & Vendy, 2002)。例如，在面对面咨询的场景之中，来访者往往能够敏感地捕捉咨询师的微表情，以评估咨询师的价值观及评判意图(Matsumoto et al., 2011)。而在 ISIs 的情境之中，若赋予虚拟化身较高的模型面数(Tris)并精细化其骨骼(Bone)设计，化身不但能模拟生动的宏表情，甚至也可能模拟出积极的微表情以进一步促进真实且融洽的交谈氛围。

4.2 倾听鼓励

咨询师对咨询工作的投入通常被认为是 TA 的预测因素，通常包含了积极地倾听与适时鼓励(朱旭, 江光荣, 2011b)。而在无指导 ISI 的研究中，研究者也发现无论被试倾诉的对象是真人还是聊天机器人，情绪宣泄所达到的效果并无显著差异(Ho et al., 2018)。不过，ISIs 程序往往需要与服务器通讯，因此以纯文本聊天机器人进行倾听与回应时，其回复速度也会影响其社交吸引力(Lew & Walther, 2022)，而基于上下文的动态回复相对于过于即时或延迟的效果更好(Samsudin, 2020)。此外，聊天机器人在动态回复时若能模拟“正在输入”状态，用户可更明显地感知到其适时的停顿与犹豫，并产生其正在“思考”的印象。对于鼓励，Chikersal 等人(2020)也指出，在 ISIs 中，那些带来更积极影响的支持信息，不但更为简短，而且也包含更多积极、肯定、鼓励等词汇。基于此，聊天机器人除了直接鼓励用户之外，在多轮对话中通过关键词复述也能给予用户间接鼓励，并提升其倾诉体验。当聊天机器人具有虚拟形象时，则可在用户表达时，针对对应的内容，回应以适当的眼神凝视、积极的面部表情、点头以及开放的手势等作为言语鼓励的补充，这能有效地使用户产生人际互动的感知(Cummins & Cui, 2014)，进而将关系线索归因于聊天机器人(Hortensius & Cross, 2018)，并进一步提升倾听、鼓励的效果。然而，鼓励可能并不适用于所有群体(Arndt et al., 2020)。因此，聊天机器人可先甄别出哪些人群更容易受鼓励的积极影响后再做反应。

4.3 真诚理解

心理咨询的效果往往取决于咨访关系的质量——若咨询师善解人意、真诚一致并无条件积极地关注着来访者，咨询效果则更好(Rogers, 1957)。而在无指导 ISI 的情境中，若聊天机器人能经常对被试的话语进行关注，并以诚实、谦虚的态度对被试进行请教，被试对它的积极评价也会更多(Zhou et al., 2019)，这不仅会让被试更具有

参与感，而且还会帮助聊天机器人“学习”新概念。此外，若聊天机器人能主动向被试披露其“个人信息”，也有可能让被试感受到它的“真诚”，进而作更多的自我暴露(Kang & Gratch, 2014)。对于理解，共情技术在传统咨询中较为常用，而在 ISIs 的情境中，聊天机器人基于真实的咨询语料来进行训练，亦可具备复述的能力来做到一定程度的“共情”。然而，复述并非鹦鹉学舌即可，而是需要用“自己的话”加上来访者话中的重要词语来提炼内容。目前，基于先进的自然语言生成模型 GPT-3(General Pre-trained Transformer-3)就可使聊天机器人做到长话短说、取其精要(Sezgin et al., 2022)。但若要推测来访者的言外之意来达到更高级的“共情”且能兼顾对话历史来保持咨询的连贯性，则仍需技术的进一步迭代。此外，聊天机器人还可基于个性化技术对用户独特的需求、偏好、情绪进行积极、精准的响应(Valentine et al., 2022)，进而使用户体验到一种区别于传统咨询的一种独特“理解”。例如，Liu-Thompkins 等人(2022)尝试将一个个性化的系统框架融入于真实的营销场景中，使聊天机器人通过偏好分析、人格评估、目标推理三个步骤，来具备换位思考的“共情”能力。有研究发现，引入个性化的设计有利于被试与 ISIs 程序在治疗任务及目标上达成一致(Penedo et al., 2020)，进而加强参与动机(Liu, 2013)及 DTA(Oinas-Kukkonen & Harjumaa, 2009; Tremain et al., 2020; Valentine et al., 2022)。

4.4 相互信赖

在一段咨访关系中，若来访者认为咨询师可信，他们才会作更多自我暴露，进而促进 TA 的发展(Bachelor, 2013; 朱旭, 江光荣, 2011b)。同样，可信度与 DTA 的质量也高度相关，ISIs 程序的低可信度可能会导致被试参与度低甚至脱落(Mackie et al., 2017)。反之，若被试觉得 ISIs 程序可信，其继续使用的意愿(Radomski et al., 2019)及其对被治愈的期望也会更高(Sauer-Zavala et al., 2018)。在无指导 ISI 中，聊天机器人所营造的拟人的第一印象会影响其可信度(Kelders et al., 2012; Neuberg, 1989; Oinas-Kukkonen & Harjumaa, 2009)，而外在刺激特征则是最关键的预测因素(Kim et al., 2021; Richards et al., 2020; Tremain et al., 2020)。但不同于传统咨询的是，在 ISIs 的情境中可让用户自主设计、搭配，或基于用户画像来赋予聊天机器人特定的种族、形象、年龄、性别、个性、声音(Brown et al., 2013)，并基于用户反馈迭代、调整，因此更具灵活性。此外，类同于复述，包含了情感词语的情感反映技术也值得探究。在干预早期，通过基于深度学习的情感预测技术(Kumar & Senthil, 2021)，聊天机器人能以简短的情感反映(如“我感到你现在很焦虑”)与用户迅速建立信任。随着 DTA 水平的逐步提升，聊天机器人还可进一步将更为关键的情感反馈给用户，并通过询问以澄清其情感体验，促使其作更深入的暴露。但由于情感通常以隐喻、明喻、举例等方式表达，因此，聊天机器人在不明其意时可灵活运用主动提问来进行确认及“学习”，以丰富知识图谱(Yin et al., 2017)。最后，并非所有用户都对情感反映表示欢迎，因此聊天机器人在作情感反映前，需综合评估个人知识图谱、DTA 水平及上下文情感词出现的频率、强度等因素来确定回复的时机及内容，并结合用户的后续反馈来习得其偏好。

综上所述，文章对现有研究结果进行了归纳，并梳理了 DTA 的前因、后果。基于此，文章将提出一个模型(见图 1)。并假设，DTA 对治疗效果有直接影响；DTA 对参与度有直接影响；参与度对治疗效果有直接影响。同时，友好尊重、倾听鼓励、真诚理解、相互信赖等关系线索或可对 DTA 造成影响，进而带来更优的参与度及治疗效果。

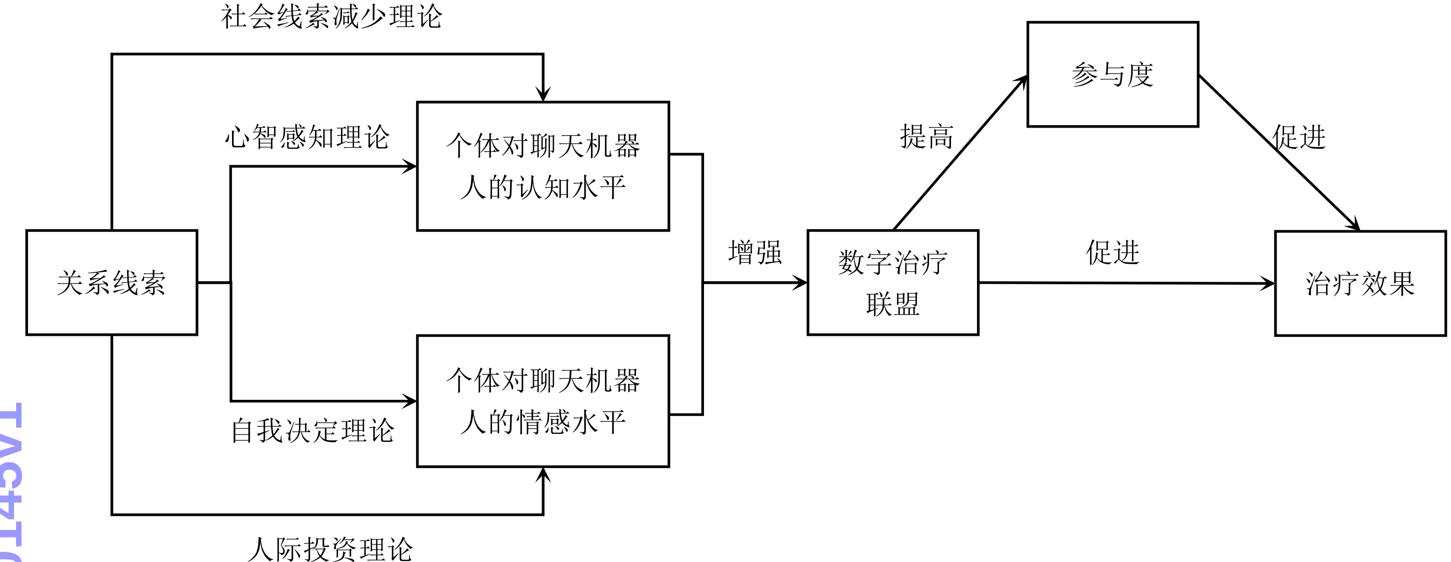


图 1　基于心智感知理论、社会线索减少理论、人际投资理论、自我决定理论视角下聊天机器人促进数字治疗联盟(DTA)的理论模型(注：箭头表示可能的因果关系方向)

注：参考 Deci & Ryan, 1985; Kiesler et al., 1984; Rusbult et al., 1994; Waytz et al., 2010 制作

5　存在问题及未来展望

5.1 需进一步探索 DTA 的影响因素

尽管人类线索的效用显著(Rietz et al., 2019)，但目前关于聊天机器人的研究多集中于言语、视觉线索，将关系线索与不同线索比较的研究相对较少(Bao et al., 2022; Grekin et al., 2019; ter Stal et al., 2020)。然而，在传统咨询领域中，研究者会将关系线索与其他变量对比，以确定不同变量对疗效的贡献大小，但文章仅涉及可能影响 DTA 的部分关系线索(Gallen et al., 2018)。因此，究竟是哪一个变量在发挥关键作用，仍不得而知(Heim et al., 2018)。在基于聊天机器人的无指导 ISI 程序中，除人类线索之外还包括用户体验、AI 对话水平、用户期望等影响因素。首先，基于 Hentati 等人(2021)的发现，程序用户界面(User Interface, UI)的易用与否对被试的参与度有显著的影响。因此，这一额外变量可能会导致研究者错误地评估聊天机器人的作用。其次，有研究发现人类线索之间存在交互作用，当聊天机器人呈现强视觉线索(人类照片)时，AI 对话水平与被试态度无关，但呈现弱视觉线索(气泡图)时，强 AI 对话水平补偿了弱视觉线索的低拟人化效果。最后，此研究还指出身份线索设定了被试对聊天机器人性能的期望，当聊天机器人被识别为人类时被试对其有更高的期望，而低 AI 对话水平则会导致更多负面评价(Go & Sundar, 2019)。因此，聊天机器人呈现的人类线索并非越多越好，不同线索的影响不同且内在关系复杂。在未来的研究中，研究者可评估更多的 TA 助长因素，并将其它人类线索及变量与之对比或组合，探索不同变量之间可能存在的交互作用。

5.2 ISIs 需作进一步的技术迭代

时下，ISIs 多是将传统心理疗法数字化，而计算机科学领域仍有诸多成果可促进 ISIs 的技术迭代。首先，可将其他成熟的结构化技术与 ISIs 程序结合，使之进一步体系化。例如，以说服性系统设计(Persuasive System Design, PSD)来构建 ISIs 程序，程序将以更多地支持、提醒、安排来提高参与度(Baumel & Yom-Tov, 2018)。此外，使用动机性访谈(Motivational Interviewing, MI)这种结构化的对话技术，也可提升用户改变的动机(Rollnick et al., 2010)，进而快速且有效地提升其参与度(Malins et al., 2020)。其次，可使用较先进的算法模型来进一步提升 ISIs 程序的性能。例如，以创新的 BERT(Bidirectional Encoder Representation from Transformers)或 GPT-3 模型来替代依赖人力、拓展性较差的 LIWC(Linguistic Inquiry and Word Count)模型(Tanana et al., 2021)。如此，聊天机器人不但能动态评估用户的情绪及 DTA 水平，其情感识别/交互能力也能得到极大的加强(Rajagopal et al., 2021)。然而，当聊天机器人的回复生成更具灵活性且富有创意时，其生成内容的不确定性也是一把双刃剑。因此，在咨询情境中将检索式与生成式结合，开发联合型的聊天机器人，或许更有利于实际的应用(Song et al., 2018)。最后，未来也可将 NLP 及计算机视觉(Computational Vision, CV)相结合的多模态技术运用于 ISIs 中。例如，通过深度学习模型对语音、语调、语速、宏表情、微表情、肢体动作、瞳孔扩张等因素进行综合分析，以进一步提升聊天机器人对用户意图、情绪推断的准确度(Hu et al., 2018; Jonell, 2019; Kuo et al., 2021; Lee et al., 2020; Liu & Yang, 2021)，并提供诸如文字、图像、选项、语音等交互形式，以适应不同群体的习惯。此外，还可结合虚拟现实(Virtual Reality, VR)技术来弥补虚拟与真实交互的差距，强化沉浸感及社会存在(Donker et al., 2019; Miloff et al,. 2020)。

5.3 开发适应 ISIs 的 DTA 测量工具并结合客观数据进行报告

具备有效的测量工具，是推进领域研究发展的重要条件，但在现阶段，研究者们对于如何衡量 DTA 却几乎没有共识(Gómez Penedo et al., 2020)。例如，研究者们要么直接使用 WAI 量表，要么只对 WAI 量表进行最小程度的微调(Ellis-Brush, 2020)，但简单地将“咨询师”替换为“应用程序”可能有失偏颇。一方面，将 ISIs 程序定位为人时，被试可能产生更高的预期并提高评价标准(Go & Sundar, 2019)。另一方面，面对面治疗中的重要因素在 ISIs 中可能并非同等重要，ISIs 具有其自身的特殊性及复杂性(Clarke et al., 2016)，而在基于聊天机器人的 ISIs 中，研究者不但要考量程序的交互体验，还需要对其中的人类线索进行评估。因此，研究者未来可在传统 TA 理论的基础上，还考虑如社会行动者范式(CASA)、恐怖谷效应(UVE)等 HCI 理论(Smelser & Baltes, 2001; Zhang et al., 2020)，针对 ISIs 情境及干预形式来设计专门的 DTA 量表(D’Alfonso et al., 2020; Heim et al., 2018)。此外，越来越多关于 TA 的研究强调，需要更准确地识别 TA 的建立以及破裂的发生(Colli et al., 2019)，但时下 DTA 的测量几乎都依赖被试的自我报告(Berger, 2017)，而没有结合行为、生理数据等进行更为客观的量化分析。因此，未来可结合更详尽的客观数据(Nof et al., 2021)，对被试的声学特征、行为轨迹、文本及视听数据进行建模，并动态分析当下的 DTA 质量，监测聊天机器人与被试的 DTA 在何时建立、破裂，进而为聊天机器人的行为决策提供更优的指导。最后，研究者还可综合评估量化数据及咨询师、观察者的主观数据，以加强研究结果的严谨性。

5.4 关注在 ISIs 中不同疗法及不同群体于 DTA 上所呈现出的新问题

目前，DTA 研究中所使用的 ISIs 程序多是基于 CBT 设计的，尽管在线 CBT 的可行、有效性均得到验证(Newby et al., 2017; Titov et al., 2015)，但仍有部分群体并

未充分受益于此(Rozental et al., 2019; Sunderland et al., 2012)。因此，在 ISIs 中仍要开发和测试更多的替代疗法。例如，正念干预(Mindfulness-based Interventions, MBIs)就被认为是 CBT 的有效替代(Li et al., 2021)。有研究表明，被试不但在 MBIs 中的 TA 得分高于 CBT(Jazaieri et al., 2018)，且状态正念也与 TA 存在高度的相关(Johnson, 2018)。但是，MBIs 与 TA 关系的研究仍然较少，在 ISIs 环境中的类似证据则更是缺乏。因此，未来的研究可在 DTA 的研究中使用正念减压疗法(Mindfulness-based Stress Reduction, MBSR)、正念认知疗法(Mindfulness-based Cognitive Therapy, MBCT)、接纳与承诺疗法(Acceptance and Commitment Therapy, ACT)等 MBIs，并尝试以聊天机器人模拟正念教练，优化现有在线 MBIs 的体验。此外，基于一种疗法的 ISIs 在不同心理问题(如抑郁、焦虑、恐惧、成瘾等)、群体(如青少年、成年人、老年人或男性、女性等)中所建立的 DTA 水平可能存在差异，但现有研究对此少有讨论(Darcy et al., 2021; Ellis-Brush, 2020; Werz et al., 2021)。因此，未来的研究在探讨 DTA 与某一症状的关系时，还可将被试划分为更多的亚组，以检验不同特征人群的结果差异，进而加深对 DTA 作用机制的理解。

参考文献

张夏恒，李想. (2022). 国外元宇宙领域研究现状、热点及启示. *产业经济评论*, *10*(2), 199 – 214.

朱旭，江光荣. (2011a). 工作同盟的概念. *中国临床心理学杂志*, *19*(2), 275 – 280.

朱旭，江光荣. (2011b). 当事人眼里的工作同盟：质的分析. *心理学报*, *43*(4), 420 – 431.

黄欣荣，曹贤平. (2022). 元宇宙的技术本质与哲学意义. *新疆师范大学学报(哲学社会科学版)*, *43*(3), 20 – 27.

Abd-alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., & Househ, M. (2019). An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics*, *132*, 103978. https://doi.org/10.1016/j.ijmedinf.2019.103978

Alkhaldi, G., Hamilton, F. L., Lau, R., Webster, R., Michie, S., & Murray, E. (2016). The effectiveness of prompts to promote engagement with digital interventions: A systematic review. *Journal of Medical Internet Research*, *18*(1), e6. https://doi.org/10.2196/jmir.4790

Ames, HM., Glenton, C., Lewin, S., Tamrat, T., Akama, E., & Leon, N. (2019). Clients' perceptions and experiences of targeted digital communication accessible via mobile devices for reproductive, maternal, newborn, child, and adolescent health: A qualitative evidence synthesis. *Cochrane Database Systematic Reviews*, *10*(10):CD013447. https://doi.org/10.1002/14651858.cd013447

Anderson, R. E. E., Spence, S. H., Donovan, C. L., March, S., Prosser, S., & Kenardy, J. (2012). Working alliance in online cognitive behavior therapy for anxiety disorders in youth: Comparison with clinic delivery and its role in predicting outcome. *Journal of Medical Internet Research*, *14*(3), e88. https://doi.org/10.2196/jmir.1848

Andersson, G., Paxling, B., Wiwe, M., Vernmark, K., Felix, C. B., Lundborg, L., … Carlbring, P. (2012). Therapeutic alliance in guided internet-delivered cognitive behavioural treatment of depression, generalized anxiety disorder and social anxiety disorder. *Behaviour Research and Therapy*, *50*(9), 544 – 550.

Arndt, A., Rubel, J., Berger, T., & Lutz, W. (2020). Outpatient and self-referred participants: Adherence to treatment components and outcome in an internet intervention targeting anxiety disorders. *Internet Interventions*, Advance online publication. https://doi.org/10.1016/j.invent.2020.100319

Asaeikheybari, G., Hooper, M. W., & Huang, M.C. (2021). A context-adaptive smoking cessation system using videos. *Smart Health*, *19*, 100148. https://doi.org/10.1016/j.smhl.2020.100148

Bachelor, A. (2013). Clients' and therapists' views of the therapeutic alliance: Similarities, differences and relationship to therapy outcome. *Clinical Psychology and Psychotherapy*, *20*(2), 118–135.

Bao, Z., Chen, J., Luo, Y., & Reani, M. (2022, July). *Convincing or odd: Anthropomorphic design cues in chatbots*. Paper presented at the meeting of the pacific asia conference on information systems, Sydney, AUS.

Baumeister, H., Reichler, L., Munzinger, M., & Lin, J. (2014). The impact of guidance on Internet-based mental health interventions: A systematic review. *Internet Interventions*, *1*(4), 205–215.

Baumel, A., & Kane, J. M. (2018). Examining predictors of real-world user engagement with self-guided eHealth interventions: Analysis of mobile apps and websites using a novel dataset. *Journal of Medical Internet Research*, *20*(12), 1–13.

Baumel, A., & Yom-Tov, E. (2018). Predicting user adherence to behavioral eHealth interventions in the real world: Examining which aspects of intervention design matter most. *Translational Behavioral Medicine*, *8*(5), 793–798.

Berger, T. (2017). The therapeutic alliance in internet interventions: A narrative review and suggestions for future research. *Psychotherapy Research*, *27*(5), 511–524.

Berger, T., Boettcher, J., & Caspar, F. (2014). Internet-based guided self-help for several anxiety disorders: A randomized controlled trial comparing a tailored with a standardized disorder-specific approach. *Psychotherapy*, *51*(2), 207–219.

Berry, K., Salter, A., Morris, R., James, S., & Bucci, S. (2018). Assessing therapeutic alliance in the context of mHealth interventions for mental health problems: Development of the mobile agnew relationship measure (mARM) questionnaire. *Journal of Medical Internet Research*, *20*(4), 1–8.

Berry, N., Lobban, F., & Bucci, S. (2019). A qualitative exploration of service user views about using digital health interventions for self-management in severe mental health problems. *BMC Psychiatry*, *19*(1), 1–13.

Bickmore, T., Gruber, A., & Picard, R. (2005). Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*, *59*(1), 21–30.

Bickmore, T. W., Pfeifer, L. M., & Jack, B. W. (2009). Taking the time to care: Empowering low health literacy hospital patients with virtual nurse agents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*(pp. 1265–1274). New York, NY: Association for Computing Machinery.

Bielinski, L. L., & Berger, T. (2020). Internet interventions for mental health: Current state of research, lessons learned and future directions. *Counseling Psychology and Psychotherapy*, *28*(3), 65–83.

Bock, B. C., Rosen, R. K., Barnett, N. P., Thind, H., Walaska, K., Foster, R., … Traficante, R. (2015). Translating behavioral interventions onto mhealth platforms: Developing text message interventions for smoking and alcohol. *JMIR MHealth and UHealth*, *3*(1), 1–9.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice*, *16*(3), 252–260.

Brown, A., Mountford, V. A., & Waller, G. (2013). Is the therapeutic alliance overvalued in the treatment of eating disorders? *International Journal of Eating Disorders*, *46*(8), 779–782.

Chen, H., Rodriguez, M. A., Qian, M., Kishimoto, T., Lin, M., & Berger, T. (2020). Predictors of treatment outcomes and adherence in internet-based cognitive behavioral therapy for social anxiety in China. *Behavioural and Cognitive Psychotherapy*, *48*(3), 291–303.

Chikersal, P., Belgrave, D., Doherty, G., Enrique, A., Palacios, J. E., Richards, D., & Thieme, A. (2020). Understanding client support strategies to improve clinical outcomes in an online mental health intervention. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*(pp. 1–16). New York, NY: Association for Computing Machinery.

Clarke, J., Proudfoot, J., Whitton, A., Birch, M.R., Boyd, M., Parker, G., … Fogarty, A. (2016). Therapeutic alliance with a fully automated mobile phone and web-based intervention: Secondary analysis of a randomized controlled trial. *JMIR Mental Health*, *3*(1), e10. https://doi.org/10.2196/mental.4656

Colli, A., Gentile, D., Condino, V., & Lingiardi, V. (2019). Assessing alliance ruptures and resolutions: Reliability and validity of the collaborative interactions scale-revised version. *Psychotherapy Research*, *29*(3), 279–292.

Cummins, R. G., & Cui, B. (2014). Reconceptualizing address in television programming: The effect of address and affective empathy on viewer experience of parasocial interaction. *Journal of Communication*, *64*(4), 723–742.

D’Alfonso, S., Lederman, R., Bucci, S., & Berry, K. (2020). The digital therapeutic alliance and human-computer interaction. *JMIR Mental Health*, *7*(12), e21895. https://doi.org/10.2196/21895

Darcy, A., Daniels, J., Salinger, D., Wicks, P., Robinson, A., & Darcy, A. (2021). Evidence of human-level bonds established with a digital conversational agent: Cross-sectional , retrospective observational study. *JMIR Formative Research*, *5*(5), 1–7.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum.

Donker, T., Cornelisz, I., Van Klaveren, C., Van Straten, A., Carlbring, P., Cuijpers, P., & Van Gelder, J. L. (2019). Effectiveness of self-guided app-based virtual reality cognitive behavior therapy for acrophobia: A randomized clinical trial. *JAMA Psychiatry*, *76*(7), 682–690.

Ellis-Brush, K. (2020). *Coaching in a digital age: can a working alliance form between coachee and “coaching app”?*(Unpublished doctorial dissertation).Oxford Brookes University.

Elmasri, D., & Maeder, A.(2016). A conversational agent for an online mental health intervention. In G. Ascoli, M. Hawrylycz, H. Ali, D. Khazanchi, & Y. Shi(Eds.), *Lecture notes in Computer Science: Vol. 9919: Brain Informatics and Health*(pp. 243–251). Berlin, Germany: Springer.

Fenski, F., Rozental, A., Heinrich, M., Knaevelsrud, C., Zagorscak, P., & Boettcher, J. (2021). Negative effects in internet-based interventions for depression: A qualitative content analysis. *Internet Interventions*, *26*, 100469. https://doi.org/10.1016/j.invent.2021.100469

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, *4*(2), 1–11.

Fuhr, K., Schröder, J., Berger, T., Moritz, S., Meyer, B., Lutz, W., Hohagen, F., … Klein, J. P. (2018). The association between adherence and outcome in an Internet intervention for depression. *Journal of Affective Disorders*, *229*, 443–449.

Gaffney, H., Mansell, W., Edwards, R., & Wright, J. (2014). Manage your life online (MYLO): A pilot trial of a conversational computer-based intervention for problem solving in a student sample. *Behavioural and Cognitive Psychotherapy*, *42*(6), 731–746.

Gallen, S., Gallen, S., Zurich, E. T. H., & Zurich, E. T. H. (2018, June). *The impact of interpersonal closeness cues in text-based healthcare chatbots on attachment bond and the desire to continue interacting: An experimental design*. Paper presented at the 26th European Conference on Information Systems, Portsmouth, UK.

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, *97*, 304–316.

Goldberg, S. B., Baldwin, S. A., Riordan, K. M., Torous, J., Dahl, C. J., Davidson, R. J., & Hirshberg, M. J. (2021). Alliance with an unguided smartphone app: Validation of the digital working alliance inventory. *Assessment*, *29*(6), 1331–1345.

Gómez Penedo, J. M., Berger, T., grosse Holtforth, M., Krieger, T., Schröder, J., Hohagen, F., … Klein, J. P. (2020). The working alliance inventory for guided internet interventions (WAI-I). *Journal of Clinical Psychology*, *76*(6), 973–986.

Grekin, E. R., Beatty, J. R., & Ondersma, S. J. (2019). Mobile health interventions: Exploring the use of common relationship factors. *JMIR MHealth and UHealth*, *7*(4), e11245. https://doi.org/10.2196/11245

Hardeman, W., Houghton, J., Lane, K., Jones, A., & Naughton, F. (2019). A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. *International Journal of Behavioral Nutrition and Physical Activity*, *16*(1), 1–21.

Hargreaves, A., Dillon, R., Castorina, M., Furey, E., Walsh, J., Fitzmaurice, B., … Donohoe, G. (2018). Predictors of adherence to low support, computerised, cognitive remediation training in psychosis. *Psychosis*, *10*(4), 298–306.

Heim, E., Rötger, A., Lorenz, N., & Maercker, A. (2018). Working alliance with an avatar: How far can we go with internet interventions? *Internet Interventions*, *11*, 41–46.

Henson, P., Wisniewski, H., Hollis, C., Keshavan, M., & Torous, J. (2019). Digital mental health apps and the therapeutic alliance: Initial review. *BJPsych Open*, *5*(1), 1–5.

Hentati, A., Forsell, E., Ljótsson, B., Kaldo, V., Lindefors, N., & Kraepelien, M. (2021). The effect of user interface on treatment engagement in a self-guided digital problem-solving intervention: A randomized controlled trial. *Internet Interventions*, *26*, 100448. https://doi.org/10.1016/j.invent.2021.100448

Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*, *68*(4), 712–733.

Holter, M. T. S., Ness, O., Johansen, A. B., & Brendryen, H. (2020). Making come-alive and keeping un-alive: How people relate to self-guided web-based health interventions. *Qualitative Health Research*, *30*(6), 927–941.

Hortensius, R., & Cross, E. S. (2018). From automata to animate beings: the scope and limits of attributing socialness to artificial agents. *Annals of the New York Academy of Sciences*, *1426*(1), 93–110.

Horton, D., & Richard Wohl, R. (1956). Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry*, *19*(3), 215–229.

Horvath, A. O., & Greenberg, L. S. (1989). Development of the working alliance inventory. *Journal of Counseling Psychology*, *36*(2), 223–233.

Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., … Akkiraju, R. (2018, April). *Touch your heart: A tone-aware chatbot for customer care on social media*. Paper presented at the 2018 CHI Conference on Human Factors in Computing Systems. Montreal, Canada.

Izzaty, R. E., Astuti, B., & Cholimah, N. (2021). Evaluation of an online cognitive behavioural therapy, mindfulness meditation, and yoga (CBT-MY) intervention for posttraumatic stress disorder: A single arm clinical trial with psychometric and psychophysiological outcomes. *Angewandte Chemie International Edition*, *6*(11), 951–952.

Jasper, K., Weise, C., Conrad, I., Andersson, G., Hiller, W., & Kleinstäuber, M. (2014). The working

alliance in a randomized controlled trial comparing internet-based self-help and face-to-face cognitive behavior therapy for chronic tinnitus. *Internet Interventions*, *1*(2), 49–57.

Jazaieri, H., Goldin, P. R., & Gross, J. J. (2018). The role of working alliance in CBT and MBSR for social anxiety disorder. *Mindfulness*, *9*(5), 1381–1389.

Johansson, M., Berman, A. H., Sinadinovic, K., Lindner, P., Hermansson, U., & Andréasson, S. (2021). Effects of internet-based cognitive behavioral therapy for harmful alcohol use and alcohol dependence as self-help or with therapist guidance: Three-armed randomized trial. *Journal of Medical Internet Research*, *23*(11), e29666. https://doi.org/10.2196/29666

Johnson, D. A. (2018). The relationship between state mindfulness and working alliance among counselors-in-training. *Journal of Humanistic Counseling*, *57*(1), 31–50.

Jonell, P. (2019). Using social and physiological signals for user adaptation in conversational agents. In N. Agmon, M. E. Taylor, E. Elkind, & M. Veloso(Eds.), *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems*(pp. 2420–2422). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Kaiser, J., Hanschmidt, F., & Kersting, A. (2021). The association between therapeutic alliance and outcome in internet-based psychological interventions: A meta-analysis. *Computers in Human Behavior*, *114*, 106512. https://doi.org/10.1016/j.chb.2020.106512

Kang, S. H., & Gratch, J. (2014). Exploring users' social responses to computer counseling interviewers' behavior. *Computers in Human Behavior*, *34*, 120–130.

Kelders, S. M., Kok, R. N., Ossebaard, H. C., & Van Gemert-Pijnen, J. E. W. C. (2012). Persuasive system design does matter: A systematic review of adherence to web-based interventions. *Journal of Medical Internet Research*, *14*(6), e152. https://doi.org/10.2196/jmir.2104

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, *39*(10), 1123–1134.

Kiluk, B. D., Serafini, K., Frankforter, T., Nich, C., & Carroll, K. M. (2014). Only connect: The working alliance in computer-based cognitive behavioral therapy. *Behaviour Research and Therapy*, *63*, 139–146.

Kim, H. C., Cha, M. C., & Ji, Y. G. (2021). The impact of an agent's voice in psychological counseling: Session evaluation and counselor rating. *Applied Sciences*, *11*(7), 2893. https://doi.org/10.3390/app11072893

Klasen, M., Knaevelsrud, C., & Böttche, M. (2013). Die therapeutische beziehung in internetbasierten therapieverfahren: Ein Überblick. *Nervenarzt*, *84*(7), 823–831.

Kumar, T. S., & Senthil, T. (2021). Construction of hybrid deep learning model for predicting children behavior based on their emotional reaction. *Journal of Information Technology*, *3*(1), 29–43.

Kuo, P. H., Lin, S. T., Hu, J., & Huang, C. J. (2021). Multi-sensor context-aware based chatbot model: An application of humanoid companion robot. *Sensors*, *21*(15), 5132.

Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., … Coiera, E. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association*, *25*(9), 1248–1258.

Lederman, R., & D'Alfonso, S. (2021). The Digital therapeutic alliance: Prospects and considerations. *JMIR Mental Health*, *8(7)*, e31385. https://doi.org/10.2196/31385

Lee, S., Lee, N., & Sah, Y. J. (2020). Perceiving a mind in a chatbot: Effect of mind perception and social cues on co-presence, closeness, and intention to use. *International Journal of Human-Computer Interaction*, *36*(10), 930–940.

Lew, Z., & Walther, J. B. (2022). Social scripts and expectancy violations: Evaluating communication with human or ai chatbot interactants. *Media Psychology, 25*(6), 1 - 16.

Li, J., Cai, Z., Li, X., Du, R., Shi, Z., Hua, Q., … Zhan, X. (2021). Mindfulness-based therapy versus cognitive behavioral therapy for people with anxiety symptoms: A systematic review and meta-analysis of random controlled trials. *Annals of Palliative Medicine, 10*(7), 7596 - 7612.

Linardon, J., Cuijpers, P., Carlbring, P., Messer, M., & Fuller-Tyszkiewicz, M. (2019). The efficacy of app-supported smartphone interventions for mental health problems: A meta-analysis of randomized controlled trials. *World Psychiatry, 18*(3), 325 - 336.

Liu-Thompkins, Y., Okazaki, S., & Li, H. (2022). Artificial empathy in marketing interactions: Bridging the human-AI gap in affective and social customer experience. *Journal of the Academy of Marketing Science, 50*, 1 - 21.

Liu, H., Peng, H., Song, X., Xu, C., & Zhang, M. (2022). Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness. *Internet Interventions*, Advance online publication. https://doi.org/10.1016/j.invent.2022.100495

Liu, X. (2013). Full-text citation analysis: A new method to enhance. *Journal of the American Society for Information Science and Technology, 64*(9), 1852 - 1863.

Liu, Y., & Yang, R. (2021). Federated learning application on depression treatment robots(DTbot). In *Proceeding of the 13th International Conference on Computer Research and Development*(pp. 121 - 124). Piscataway, NJ: IEEE.

Luborsky, L. L. (1976). Helping alliances in psychotherapy. In J. L. Cleghhorn(Ed.), *Successful Psychotherapy*(pp. 92 - 116). New York, NY: Brunner/Mazel.

Luo, B., Lau, R. Y. K., Li, C., & Si, Y. (2022). A critical review of state - of - the - art chatbot designs and applications. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 12*(1), e1434. https://doi.org/10.1002/widm.1434

Mackie, C., Dunn, N., Maclean, S., Testa, V., Heisel, M., & Hatcher, S. (2017). A qualitative study of a blended therapy using problem solving therapy with a customised smartphone app in men who present to hospital with intentional self-harm. *Evidence-Based Mental Health, 20*(4), 118 - 122.

Malins, S., Biswas, S., Rathbone, J., Vogt, W., Pye, N., Levene, J., … Russell, J. (2020). Reducing dropout in acceptance and commitment therapy, mindfulness-based cognitive therapy, and problem-solving therapy for chronic pain and cancer patients using motivational interviewing. *British Journal of Clinical Psychology, 59*(3), 424 - 438.

Matsumoto, D., Hwang, H. S., Skinner, L., & Frank, M. (2011). Evaluating truthfulness and detecting deception. *FBI Law Enforcement Bulletin, 6*(80), 1 - 8.

Medhi Thies, I., Menon, N., Magapu, S., Subramony, M., & O' neill, J. (2017). How do you want your chatbot? An exploratory Wizard-of-Oz study with young, urban Indians. In R. Bernaupt, G. Dalvi, A. Joshi, K. Balkrishan, J. O'Neill, & M. Winckler(Eds.), *Proceeding of the IFIP Conference on Human-Computer Interaction*(pp. 441 - 459). Berlin, Germany: Springer.

Meyer, B., Bierbrodt, J., Schröder, J., Berger, T., Beevers, C. G., Weiss, M., … Klein, J. P. (2015). Effects of an Internet intervention (Deprexis) on severe depression symptoms: Randomized controlled trial. *Internet Interventions, 2*(1), 48 - 59.

Miloff, A., Carlbring, P., Hamilton, W., Andersson, G., Reuterskiöld, L., & Lindner, P. (2020). Measuring alliance toward embodied virtual therapists in the era of automated treatments with the Virtual Therapist Alliance Scale (VTAS): Development and psychometric evaluation. *Journal of Medical Internet Research, 22*(3), 1 - 13.

Mu, U. (2021). Digital encounters: Human interactions in mHealth behavior change interventions.

Digital Health, *7*, 1 - 6.

Nass, C., Steuer, J., & Tauber, E. R. (1994, April). *Computers are social actors*. Paper presented at *the SIGCHI Conference on Human Factors in Computing Systems*(pp. 72 - 78), Boston, MA.

Negri, A., Christian, C., Mariani, R., Belotti, L., Andreoli, G., & Danskin, K. (2019). Linguistic features of the therapeutic alliance in the first session: A psychotherapy process study. *Research in Psychotherapy: Psychopathology, Process and Outcome*, *22*(1), 71 - 82.

Neuberg, S. L. (1989). The goal of forming accurate impressions during social interactions: Attenuating the impact of negative expectancies. *Journal of Personality and Social Psychology*, *56*(3), 374.

Newby, J. M., Mewton, L., & Andrews, G. (2017). Transdiagnostic versus disorder-specific internet-delivered cognitive behaviour therapy for anxiety and depression in primary care. *Journal of Anxiety Disorders*, *46*, 25 - 34.

Nof, A., Amir, O., Goldstein, P., & Zilcha-Mano, S. (2021). What do these sounds tell us about the therapeutic alliance: Acoustic markers as predictors of alliance. *Clinical Psychology and Psychotherapy*, *28*(4), 807 - 817.

Nofz, M. P., & Vendy, P. (2002). When computers say it with feeling: Communication and synthetic emotions in Kubrick' s 2001: A space odyssey. *Journal of Communication Inquiry*, *26*(1), 26 - 45.

Norcross, J. C. (2002). *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients.* Oxford University Press.

Oinas-Kukkonen, H., & Harjumaa, M. (2009). Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems*, *24*(1), 485 - 500.

Penedo, J. M. G., Babl, A. M., Holtforth, M. grosse., Hohagen, F., Krieger, T., Lutz, W., … Berger, T. (2020). The association of therapeutic alliance with long-term outcome in a guided internet intervention for depression: Secondary analysis from a randomized control trial. *Journal of Medical Internet Research*, *22*(3), 1 - 11.

Perski, O., Blandford, A., West, R., & Michie, S. (2017). Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Translational Behavioral Medicine*, *7*(2), 254 - 267.

Perski, O., Crane, D., Beard, E., & Brown, J. (2019). Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. *Digit Health*, *5*, 1 - 13.

Pihlaja, S., Stenberg, J. H., Joutsenniemi, K., Mehik, H., Ritola, V., & Joffe, G. (2018). Therapeutic alliance in guided internet therapy programs for depression and anxiety disorders – A systematic review. *Internet Interventions*, *11*, 1 - 10.

Pratap, A., Neto, E. C., Snyder, P., Stepnowsky, C., Elhadad, N., Grant, D., … Omberg, L. (2020). Indicators of retention in remote digital health studies: A cross-study evaluation of 100,000 participants. *Npj Digital Medicine*, *3*(1), 1 - 10.

Probst, G. H., Berger, T., & Flückiger, C. (2019). The alliance-outcome relation in internet-based interventions for psychological disorders: A correlational meta-analysis. *Verhaltenstherapie*, *29*(3), 182 - 195.

Provoost, S. (2021). *Embodied Conversational Agents in Internet-Based Cognitive Behavioral Therapy for Depression Bridging the Gap Between Unguided and Guided Interventions*(Unpublished doctorial dissertation). Vrije Universiteit Amsterdam.

Provoost, S., Lau, H. M., Ruwaard, J., & Riper, H. (2017). Embodied conversational agents in clinical psychology: A scoping review. *Journal of Medical Internet Research*, *19*(5), e151.

https://doi.org/10.2196/jmir.6553
Puls, H. C., Schmidt, R., Herpertz, S., Zipfel, S., Tuschen-Caffier, B., Friederich, H. C., … Hilbert, A. (2020). Adherence as a predictor of dropout in internet-based guided self-help for adults with binge-eating disorder and overweight or obesity. *International Journal of Eating Disorders*, *53*(4), 555 – 563.
Radomski, A. D., Wozney, L., McGrath, P., Huguet, A., Hartling, L., Dyson, M. P., … Newton, A. S. (2019). Design and delivery features that may improve the use of internet-based cognitive behavioral therapy for children and adolescents with anxiety: A realist synthesis with a persuasive systems design perspective. *Journal of Medical Internet Research*, *21*(2), 1 – 21.
Rajagopal, A, Nirmala, V., Andrew, J., & Arun, M. (2021). Novel AI to avert the mental health crisis in COVID- 19 : Novel application of GPT2 in cognitive behaviour therapy. *Research Square*, Advance online publication. https://doi.org/https://doi.org/10.21203/rs.3.rs-382748/v3
Richards, D., Alsharbi, B., & Abdulrahman, A. (2020). Can I help you? Preferences of young adults for the age, gender and ethnicity of a virtual support person based on individual differences including personality and psychological state. In *Proceedings of the Australasian Computer Science Week Multiconference*(pp. 1 – 10). New York, NY: Association for Computing Machinery.
Rietz, T., Benke, I., & Maedche, A. (2019). The impact of anthropomorphic and functional chatbot design features in enterprise collaboration systems on user acceptance. In *Proceedings of the 14th International Conference on Wirtschaftsinformatik*(pp. 1642 – 1656).
Rodrigues, T. R. C., Reijnders, T., Buisonjé, D. R. De., Kowatsch, T., Janssen, V. R., Kraaijenhagen, R. A., … Evers, A. W. M. (2021). Human cues in self-help lifestyle interventions : An experimental field study table of contents. *Journal of Medical Internet Research*, Advance online publication. https://doi.org/10.2196/preprints.30057
Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology*, *21*(2), 95 – 103.
Rollnick, S., Butler, C. C., Kinnersley, P., Gregory, J., & Mash, B. (2010). Motivational interviewing. *BMJ*, *340*, 1242 – 1245.
Rozental, A., Andersson, G., & Carlbring, P. (2019). In the absence of effects: An individual patient data meta-analysis of non-response and its predictors in internet-based cognitive behavior therapy. *Frontiers in Psychology*, *10*, 589.
Rusbult, C. E., Drigotas, S. M., & Verette, J. (1994). The investment model: An interdependence analysis of commitment processes and relationship maintenance phenomena. In D. J. Canary & L. Stafford (Eds.), *Communication and relational maintenance*(pp. 115 – 139). Academic Press.
Samsudin, C. M. (2020). *The effect of chatbots response latency on users’ trust* (Unpublished doctorial dissertation). University of Oklahoma.
Sauer-Zavala, S., Boswell, J. F., Bentley, K. H., Thompson-Hollands, J., Farchione, T. J., & Barlow, D. H. (2018). Expectancies, working alliance, and outcome in transdiagnostic and single diagnosis treatment for anxiety disorders: An investigation of mediation. *Cognitive Therapy and Research*, *42*(2), 135 – 145.
Sezgin, E., Sirrianni, J., & Linwood, S. L. (2022). Operationalizing and implementing pretrained, large artificial intelligence linguistic models in the US health care system: Outlook of generative pretrained transformer 3 (GPT-3) as a service model. *JMIR Medical Informatics*, *10*(2), 1 – 7.
Smelser, N. J., & Baltes, P. B.(Eds). (2001). *International encyclopedia of the social & behavioral sciences* (Vol. 11). Amsterdam, Netherlands: Elsevier.

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & Behavior*, *7*(3), 321 - 326.

Song, Y., Yan, R., Li, C.-T., Nie, J.-Y., Zhang, M., & Zhao, D. (2018). An ensembel of retrieval-based and generation based human-computer conversation systems. In *Proceedings of the International Conference of Learning Representation ICLR 2018*(pp. 4382 - 4388). ICLR.

Sun, S., Lin, D., Goldberg, S., Shen, Z., Chen, P., Qiao, S., … Operario, D. (2021). A mindfulness-based mobile health (mHealth) intervention among psychologically distressed university students in quarantine during the COVID-19 pandemic: A randomized controlled trial. *Journal of Counseling Psychology*, *69*(2), 157 - 171.

Sundar, S. S. (2008). *The MAIN model: A heuristic approach to understanding technology effects on credibility*. In M. J. Metzger, A. J. Flanagin, D. John, & C. T. MacArthur(Eds.), *Digital Media, Youth, and Credibility*(pp. 73 - 100). Cambridge, MA: MIT press.

Sunderland, M., Wong, N., Hilvert-Bruce, Z., & Andrews, G. (2012). Investigating trajectories of change in psychological distress amongst patients with depression and generalised anxiety disorder treated with internet cognitive behavioural therapy. *Behaviour Research and Therapy*, *50*(6), 374 - 380.

Tanana, M. J., Soma, C. S., Kuo, P. B., Bertagnolli, N. M., Dembe, A., Pace, B. T., … Imel, Z. E. (2021). How do you feel? Using natural language processing to automatically rate emotion in psychotherapy. *Behavior Research Methods*, *53*(5), 1 - 14.

Taylor, H., Strauss, C., & Cavanagh, K. (2021). Can a little bit of mindfulness do you good? A systematic review and meta-analyses of unguided mindfulness-based self-help interventions. *Clinical Psychology Review*, *89*, 102078. https://doi.org/10.1016/j.cpr.2021.102078

ter Stal, S., Kramer, L. L., Tabak, M., op den Akker, H., & Hermens, H. (2020). Design features of embodied conversational agents in eHealth: A literature review. *International Journal of Human Computer Studies*, *138*, 102409. https://doi.org/10.1016/j.ijhcs.2020.102409

Titov, N., Dear, B. F., Staples, L. G., Terides, M. D., Karin, E., Sheehan, J., …Wootton, B. M. (2015). Disorder-specific versus transdiagnostic and clinician-guided versus self-guided treatment for major depressive disorder and comorbid anxiety disorders: a randomized controlled trial. *Journal of Anxiety Disorders*, *35*, 88 - 102.

Tremain, H., McEnery, C., Fletcher, K., & Murray, G. (2020). The therapeutic alliance in digital mental health interventions for serious mental illnesses: Narrative review. *JMIR Mental Health*, *7*(8), e17204. https://doi.org/10.2196/17204

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *The Canadian Journal of Psychiatry*, *64*(7), 456 - 464.

Valentine, L., D'Alfonso, S., & Lederman, R. (2022). Recommender systems for mental health apps: Advantages and ethical challenges. *AI & SOCIETY*, Advance online publication. https://doi.org/10.1007/s00146-021-01322-w

Wanser, S. H., Vitale, K. R., Thielke, L. E., Brubaker, L., & Udell, M. A. R. (2019). Spotlight on the psychological basis of childhood pet attachment and its implications. *Psychology Research and Behavior Management*, *12*, 469.

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, *14*(8), 383 - 388.

Weisel, K. K., Fuhrmann, L. M., Berking, M., Baumeister, H., Cuijpers, P., & Ebert, D. D. (2019). Standalone smartphone apps for mental health: A systematic review and meta-analysis. *Npj Digital Medicine*, *2*(1), 1 - 10.

Weizenbaum, J. (1983). ELIZA: A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *26*(1), 23–28.

Werz, J., Voderholzer, U., & Tuschen-Caffier, B. (2021). Alliance matters: but how much? A systematic review on therapeutic alliance and outcome in patients with anorexia nervosa and bulimia nervosa. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity*, Advance online publication. https://doi.org/10.1007/s40519-021-01281-7

Wiese, E., Weis, P. P., Bigman, Y., Kapsaskis, K., & Gray, K. (2022). It's a match: Task assignment in human–robot collaboration depends on mind perception. *International Journal of Social Robotics*, *14*(1), 141–148.

Yao, H., Chen, J. H., & Xu, Y. F. (2020). Rethinking online mental health services in China during the COVID-19 epidemic. *Asian Journal of Psychiatry*, Advance online publication. https://doi.org/10.1016/j.ajp.2020.102015

Yin, Z., Chang, K., & Zhang, R. (2017). Deepprobe: Information directed sequence understanding and chatbot design via recurrent neural networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*(pp. 2131–2139). New York, NY: Association for Computing Machinery.

Zhang, J., Oh, Y. J., Lange, P., Yu, Z., & Fukuoka, Y. (2020). Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: Viewpoint. *Journal of Medical Internet Research*, *22*(9), e22845. https://doi.org/10.2196/22845

Zhang, X., Lewis, S., Firth, J., Chen, X., & Bucci, S. (2021). Digital mental health in China: A systematic review. *Psychological Medicine*, *51*(15), 2552–2570.

Zhou, M. X., Wang, C., Mark, G., Yang, H., & Xu, K. (2019). Building real-world chatbot interviewers: Lessons from a wizard-of-oz field study. In C. Trattner, D. Parra, & N. Riche(Eds.), *Proceedings of ACM IUI 2019 Workshops*(Vol. 2327, pp. 1–6). New York, NY: ACM.

作者贡献声明：

磨然：提出命题、设计，文献检索、整理、总结，论文起草、最终版本修订；

方作之：重要观点提出，文献检索、整理，论文最终版本修订；

方建东：论文最终版本修订。